

C0

(19)日本国特許庁(JP)

(12) 公開特許公報(A)

(11)特許出願公開番号

特開平5-224687

(43)公開日 平成5年(1993)9月3日

(51)Int.Cl.⁵

G 1 0 L 3/00

G 0 6 F 15/20

識別記号

庁内整理番号

F I

技術表示箇所

E 8948-5H

5 6 8 Z 7343-5L

審査請求 未請求 請求項の数2(全12頁)

(21)出願番号 特願平4-30232

(22)出願日 平成4年(1992)2月18日

(71)出願人 000004226

日本電信電話株式会社

東京都千代田区内幸町一丁目1番6号

(72)発明者 高木 伸一郎

東京都千代田区内幸町1丁目1番6号 日

本電信電話株式会社内

(72)発明者 中田 寿

東京都千代田区内幸町1丁目1番6号 日

本電信電話株式会社内

(72)発明者 勝俣 雅司

東京都千代田区内幸町1丁目1番6号 日

本電信電話株式会社内

(74)代理人 弁理士 森田 寛

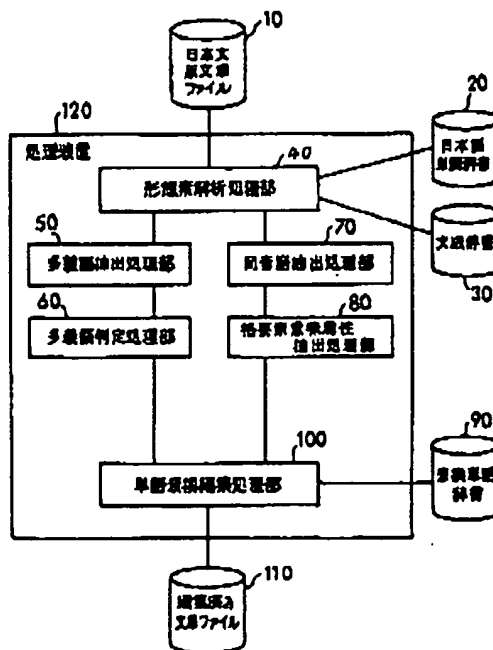
(54)【発明の名称】 日本文読み上げ単語変換編集処理方式

(57)【要約】

【目的】 本発明は、日本文文章を形態素解析処理の結果から変換すべき条件に合う同音語や多義語について単語変換辞書を検索して該当する変換単語に変換し編集するようにすることを目的としている。

【構成】 日本語単語辞書の単語情報に同音語や多義語を示す情報フラグを予め設定し、置換する変換単語を収録した変換単語辞書を備える構成としている。

本発明の処理ブロック構成例



(2)

特開平5-224687

1

【特許請求の範囲】

【請求項1】 読み上げ用の日本文文章ファイルに含まれる同音語や多義語を聞き取りやすい単語に変換し編集する処理において、

読み上げ用の日本文原文章を日本語単語辞書と文法辞書とを用いて形態素解析処理して単語の認定と単語の言語情報とを取得する手段と、

予め日本語単語辞書の単語情報に多義語を示す情報フラグを設定する手段と、

該情報フラグによって多義語を抽出する手段と、

多義語で文法的な接続条件を満たす単語候補が複数存在する場合に、文章ファイルに付随する文章分野属性と多義語の意味属性との共起性で多義を絞り込む手段と、

予め多義語の各多義毎に多義語の単語見出しと品詞に対して置換する変換単語とを収録した変換単語辞書と、

該変換単語辞書を検索して多義語を変換編集する手段とを備えることを特徴とする日本文読み上げ単語変換編集処理方式。

【請求項2】 予め日本語単語辞書の単語情報に動詞の同音語を示す情報フラグを設定する手段と、

該情報フラグによって同音語を抽出する手段と、

同音語の前方に存在する格構造関係の名詞の意味属性を抽出する手段と、

予め同音語見出しと格構造関係の名詞の意味属性との組み合わせ毎に置換する変換単語を収録した変換単語辞書と、

該変換単語辞書を検索して同音語を変換編集する手段とを備えることを特徴とする請求項1記載の日本文読み上げ単語変換編集処理方式。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、日本文のニュース速報や各種情報案内などを読み上げて通信手段で配信する際に、日本文文章の単語を聞き取りやすく変換編集する日本文読み上げ単語変換編集処理方式に関するものである。

【0002】

【従来の技術】日本文のニュース速報や各種情報案内などの即時性の必要な情報データが通信手段で伝送され電話やファクシミリなどのメディアで利用者に配信するサービスが急増している。特に日本文文章を読み上げて電話などの音声データで配信する場合には読み上げ処理が一過性であるため、読み上げ内容の聞き取りやすさが重要である。しかし、日本語には、同音語や同形で多義を有する単語があり、これが混在することによっては聞き取りや聞きにくさなどの了解度の低下が発生する。

【0003】これに対して、情報の了解度を上げるために、

①利用者は繰り返し読み上げなどの操作を行う、

②情報配信元で単語読み分け用の特殊読み付与や単語置

2

換を行う、

③情報配信元で単語表現の変換の編集作業を行う、などの対応が可能である。

【0004】

【発明が解決しようとする課題】しかし、①では、利用者の操作が増えたりスピーカなどで複数利用者に音声データを流す場合は本質的に繰り返し読み上げなどは不可能であるなどの問題点がある。

【0005】また、②では、文章校正処理の例で行われているように、原文文字をイメージしやすいような特殊読みを行う。例えば、「追及」「追究」「追求」は、「ツイキウオヨビ」「ツイキウキワメ」「ツイキウモトメ」となる。しかし、一般の情報利用者には特殊読みは聞きにくく自然性を損なってわかりにくくなる。また出現する単語をわかりやすい単語に文字的に置換する手段では、「日中間で経済問題を日中協議する」の場合で、「日中」→「日本と中国」とすると、「日本と中国間で経済問題を日本と中国協議する」となる。さらに、「汚染土壌を採る法的手段を採る」の場合で、「採る」→「採取する」とすると、「汚染土壌を採取する法的手段を採取する」となり、いずれも不適当な表現に置き換えられてしまう問題がある。

【0006】また、③では、日本文文章の編集作業段階で多数の同音語や多義語に対する変換や編集を必要な範囲で行わねばならないので熟練を要する作業であった。このように、日本文文章の読み上げ処理において聞き取りにくい同音語や多義語を抽出して聞き取りが少なくわかりやすい単語表現に変換するなどの編集を必要な範囲で行うことは、編集者の能力を有するので短時間で処理することは困難であるという問題点があった。

【0007】本発明は、上記の従来手段における問題点を解決するために、日本文文章を形態素解析処理の結果から変換すべき条件に合う同音語や多義語について単語変換辞書を検索して該当する変換単語に変換し編集するようにすることを目的としている。

【0008】

【課題を解決するための手段】上記の目的を実現するために、本発明では、読み上げ用の日本文文章ファイルに含まれる同音語や多義語を聞き取りやすい単語に変換し編集する処理において、読み上げ用の日本文原文章を日本語単語辞書と文法辞書とを用いて形態素解析処理して単語の認定と単語の言語情報とを取得する手段と、予め日本語単語辞書の単語情報に多義語を示す情報フラグを設定する手段と、該情報フラグによって多義語を抽出する手段と、多義語で文法的な接続条件を満たす単語候補（意味的多義）が複数存在する場合に、文章ファイルに付随する文章分野属性と多義語の意味属性（単語の意味を一般的な用語で示したもの）との共起性（分野依存度）で多義を絞り込む手段と、予め多義語の各多義毎に多義語の単語見出しと品詞に対して置換する変換単語と

(3)

特開平5-224687

3

を収録した変換単語辞書と、該変換単語辞書を検索して多義語を変換編集する手段と、をそなえることを特徴とし、また予め日本語単語辞書の単語情報に動詞の同音語を示す情報フラグを設定する手段と、該情報フラグによって同音語を抽出する手段と、同音語の前方に存在する格構造関係の名詞（以下、格要素単語という）の意味属性を抽出する手段と、予め同音語見出しと格要素単語の意味属性との組み合わせ毎に置換する変換単語を収録した変換単語辞書と、該変換単語辞書を検索して同音語を変換編集する手段とを備えることを特徴とする。

【0009】

【作用】本発明においては、読み上げ用の日本文文章ファイルに含まれる同音語や多義語を聞き取りやすい単語に変換し編集する処理において、日本語単語辞書の単語情報に同音語や多義語を示す情報フラグを予め設定し、置換する変換単語を収録した変換単語辞書を備えてあり、これにより、留意すべき同音語や多義語を抽出する処理や前方の格要素単語の意味属性による聞き取りやすい単語の置換処理を行うことが可能となるので、編集者の文章編集の処理量を軽減できる。

【0010】

【実施例】以下、本発明の実施例を図面により詳細に説明する。図1から図8は、本発明の一実施例を示す図である。そして、図4から図6は図1のブロック構成における本発明の請求項1の実施例を示す図であり、図7から図8は図1のブロック構成における本発明の請求項2の実施例を示す図である。

【0011】図1は処理ブロック構成例、図2は日本語単語辞書の構成例、図3は変換単語辞書の構成例、図4と図7は処理概略フロー、図5、図6、図8は単語変換編集処理の実施例を説明する図である。

【0012】図1において、処理装置120はCPUおよびメモリからなる処理装置で以下の機能部を有する。すなわち、読み上げ用の日本文原文章ファイル10の日本文文章を日本語単語辞書20、文法辞書30を用いて単語の認定や単語の言語情報の取得を行う形態素解析処理部40；単語情報から抽出した情報フラグのうち「同形多義フラグ」がオンの単語を抽出する多義語抽出処理部50；文法的な接続条件を満たす単語候補（意味的多義）が存在する場合に多義を絞り込む多義語判定処理部60；単語情報から抽出した情報フラグのうち「同音フラグ」がオンの単語を抽出する同音語抽出処理部70；同音語の前方に存在する格要素単語の意味属性を抽出する格要素意味属性抽出処理部80；多義語単語見出しと多義語置換用の変換単語ならびに同音語単語見出しと格要素単語の意味属性との組み合わせ毎に同音語置換用の変換単語を予め収録した変換単語辞書90；該変換単語辞書を検索して多義語や同音語を変換編集する単語変換編集処理部100；編集済みの文章ファイル110からなる。

4

【0013】この処理装置120では、読み上げ用の日本文原文章の任意の文章について先頭から形態素解析処理を行い、単語の認定と単語の言語情報として品詞、読み、同音フラグ、同形多義フラグ、意味属性などの認定とを行う。次に、認定された単語列の中から多義語（「同形多義フラグ」がオンの単語）を抽出する（多義語抽出処理部50）。文法的な接続条件を満たす単語候補（意味的多義）が複数存在する場合に文章ファイルに付随する文章分野属性と多義語の意味属性との共起性（分野依存度）で多義を絞り込み単語候補を決定する（多義語判定処理部60）。さらに、多義語の単語見出しと認定品詞で変換単語辞書90を検索して変換単語を抽出し原文章ファイルの文字列に置換して編集済みの文章ファイル110に書き込む（単語変換編集処理部100）。

【0014】また、認定された単語列の中から品詞が動詞の同音語（「同音フラグ」がオンの単語）を抽出し（同音語抽出処理部70）、同音語の前方に存在する格要素単語の意味属性を抽出する（格要素意味属性抽出処理部80）。

【0015】予め単語見出しと格要素単語の意味属性との組み合わせ毎に置換する変換単語を収録した変換単語辞書90を、同音語の単語見出しと抽出した格要素の意味属性で検索してマッチするパターンの変換単語を抽出し、原文章ファイルの文字列に置換して編集済みの文章ファイル110に書き込む（単語変換編集処理部100）。

【0016】図2は、図1のブロック図の構成要素である日本語単語辞書20の構成例を示す図である。「日中」では、時詞と固有名詞との2つの多義を有する単語候補があるので、同形多義フラグを付与されている。さらに、単語候補の多義を判定するために意味属性（単語の意味を一般的な用語で示したもの）として、「時間」と「国、地域」がある。また、動詞の同音語の「取る」「採る」「撮る」にはいずれも同音フラグが付与されている。

【0017】図3は、図1のブロック図の構成要素である変換単語辞書90の構成例を示す図である。ここで、130は同音語の前方の格要素単語の意味属性、140は変換単語である。「日中」では、時詞と固有名詞との2つの多義を有する単語候補があるので、それぞれの変換単語は、時詞の場合に「昼間の間」、固有名詞の場合に「日本と中国」となり、格納されている。また、動詞の同音語の場合には格要素単語の意味属性で変換単語が異なる。例えば、「採る」では、格要素単語の意味属性が「制度」の時には「採用する」、「生物」の時には「採取する」となる。

【0018】図4は図1に示した処理ブロック構成例において、日本文原文章ファイル10に含まれる多義語を変換する処理の概略フローを示す図であり、概略フロー

(4)

特開平5-224687

5

6

に従って、動作の説明を行う。

* * 【0019】

日本文原文章ファイル10より単語変換編集処理を施す処理対象文章を読み込む
(ステップ100)

読み込んだ全文章について形態素解析処理部40において、日本語単語辞書200、文法辞書30を用いて、形態素解析処理を行い、単語の認定と、単語の言語情報として品詞、読み、同音語フラグ、同形多義フラグ、意味属性などの認定とを行う
(ステップ110)

多義語抽出処理部50において、認定された単語列の中から多義語を示す情報フラグによって多義語を抽出し、多義語でない場合にはステップ180に分岐する
(ステップ120)

抽出した多義語において、文法的な接続条件を満たす単語候補(意味的多義)が複数存在するかを判定して、意味的多義がなければステップ160に分岐する
(ステップ130)

多義語判定処理部60において、意味的多義を有する多義語の場合(ステップ130の判定がYESの場合)には、文章分野属性と多義語の意味属性との共起性(分野依存度)を調べる
(ステップ140)

共起性(分野依存度)の高い方の単語候補を認定する
(ステップ150)

認定した多義語の単語見出しと認定品詞で、変換単語辞書90を検索して変換単語を抽出し原文章ファイルの文字列を置換する
(ステップ160)

編集済みの文章ファイル110に書き込む
(ステップ170)

全単語について処理を終了したかを判定して判定がYESの場合は処理を終える。また、判定がNOの場合、次単語を読み込み(ステップ190)、ステップ120へ移行し処理を継続する
(ステップ180)

図5は図4の多義語を変換する処理の概略フローにおける多義判定処理の実施例である。

【0020】ここで、150は文章分野属性、160は原文章文字列、170は多義語、180は文法的単語接続条件、190は分野依存度判定条件、195は多義判定後の認定品詞である。

【0021】抽出した多義語「日中」において、文法的な接続条件を満たす単語候補(意味的多義)が複数存在するかを判定する。「日中開かれた」では、後続単語の動詞「開かれ」との文法的接続条件で「日中(時詞)」のみが認定されるため意味的多義はない。しかし、「日中間の懸案」では、後続単語の接尾辞「間」との文法的接続条件で「日中(時詞)」と「日中(固有名詞)」との複数の単語候補が認定されるため意味的多義が発生する。意味的多義が残留すると変換単語の検索処理で支障があるため、文章分野属性150を用いて多義語の意味属性との共起性(分野依存度)を調べ多義を絞る。ここでは、文章分野属性が「国際政治」であるので、意味属性が「国、地域」の「日中(固有名詞)」の方が認定される(認定後の品詞195)。

【0022】図6は図4の多義語を変換する処理の概略フローの実施例である。ここで、200は変換単語、210は変換後文字列である。意味的多義を絞った多義語※

同音語抽出処理部70において、認定された単語列の中から動詞の同音語を示す情報フラグによって同音語を抽出し、同音語でない場合にはステップ280に分岐する
(ステップ220)

格要素意味属性抽出処理部80において、抽出した同音語について同音語の前

※に対して、認定された単語見出しと品詞で変換単語辞書90を検索して変換単語200を抽出し、原文章ファイルの文字列と置換して変換後文字列210を作成する。実施例では、「日中(時詞)」に対して「昼間の間」を変換し、「日中(固有名詞)」に対しては「日本と中国」を変換している。

【0023】このように、予め多義語の各多義毎に多義語の単語見出しと品詞と置換する変換単語を収録した変換単語辞書として作成しておき、多義語が抽出された場合に、意味的多義を絞って認定した多義語の単語見出しと品詞とで該変換単語辞書を検索して聞き取りやすい変換単語への変換を行うのであるから、編集者の文章編集の処理量を軽減できる。

【0024】図7は図1に示した処理ブロック構成例において、日本文原文章ファイル10に含まれる動詞の同音語を変換する処理の概略フローを示す図であり、概略フローに従って、動作の説明を行う。

【0025】ここで、ステップ200、ステップ210、ステップ280、ステップ290はそれぞれ、図4のステップ100、ステップ110、ステップ180、ステップ190に等しい。

【0026】

(5)

特開平5-224687

7

8

方の単語を探索して格要素単語を抽出し、格要素単語の意味属性を抽出する

(ステップ230)

同音語の単語見出しと抽出した格要素単語の意味属性で、変換単語辞書90を
検索する

(ステップ240)

変換単語辞書でマッチするパターンがあるかを判定する。マッチするパターン
がない場合にはステップ280に分岐する

(ステップ250)

変換単語を抽出し原文ファイルの文字列を置換する

(ステップ260)

編集済みの文章ファイル110に書き込む

(ステップ270)

図8は図7の同音語を変換する処理の概略フローの実施例である。

【0027】ここで、220は原文文章文字列、230は同音語(動詞)、240は格要素単語との格構造関係、250は変換単語、260は変換後文字列である。抽出した同音語「捕る」「採る」について、同音語の前方の単語を探索して格要素単語を抽出し、格要素単語の意味属性を抽出する。実施例では、「捕る」に対して『サケ(生物)』、「採る」に対して『病原菌(生物)』と『手段(制度)』がそれぞれ抽出される。次に、格要素単語を有する同音語の単語見出しと抽出した格要素単語の意味属性で、変換単語辞書90を検索し、変換単語辞書でマッチするパターンがあるかを判定する。マッチするパターンがあれば変換単語250を抽出し、原文ファイルの文字列と置換して変換後文字列260を作成する。実施例では、「捕る」に対して『サケ(生物)』が格構造関係にあるので、「捕る」を「捕獲する」に変換している。また、「採る」に対しては、格要素単語が『病原菌(生物)』に対して「採る」を「採取する」に変換し、格要素単語が『手段(制度)』に対しては「採る」を「採用する」に変換している。

【0028】このように、予め同音語見出しと格要素単語の意味属性との組み合わせ毎に置換する変換単語を収録した変換単語辞書を作成しておき、同音語が抽出された場合に、格要素単語を有する同音語の単語見出しと抽出した格要素単語の意味属性で、該変換単語辞書を検索して聞き取りやすい変換単語への変換を行う。

【0029】すなわち、本発明の日本語読み上げ単語変換編集処理方式では、読み上げ用の日本語文章ファイルに含まれる同音語や多義語を聞き取りやすい単語に変換し編集する処理において、既に記述した手段によって、留意すべき同音語や多義語を抽出する処理や前方の格要素単語の意味属性による聞き取りやすい変換単語への置換処理を行うことが可能となる。

【0030】

【発明の効果】以上説明したように、本発明によれば、特に日本語文章を読み上げて電話などの音声データで速報記事などを配信する際には、読み上げ内容の聞き取りや聞きにくさなどを排除して了解度の高い日本語文章を作成するために、一般に日本語文章の編集作業を行うに当って、編集者の文章編集の処理量を軽減できる。

*【図面の簡単な説明】

10 【図1】本発明の処理ブロック構成例である。

【図2】日本語単語辞書の構成例である。

【図3】変換単語辞書の構成例である。

【図4】処理概略フローである。

【図5】単語変換編集処理の実施例を説明する図である。

【図6】単語変換編集処理の実施例を説明する図である。

【図7】処理概略フローである。

【図8】単語変換編集処理の実施例を説明する図である。

【符号の説明】

10 日本文原文ファイル

20 日本語単語辞書

30 文法辞書

40 形態素解析処理部

50 多義語抽出処理部

60 多義語判定処理部

70 同音語抽出処理部

80 格要素意味属性抽出処理部

30 90 変換単語辞書

100 単語変換編集処理部

110 編集済みの文章ファイル

120 処理装置

130 格要素単語の意味属性

140 変換単語

150 文章分野属性

160 原文文章文字列

170 多義語

180 文法的単語接続条件

40 190 分野依存度判定条件

195 認定品詞

200 変換単語

210 変換後文字列

220 原文文章文字列

230 同音語

240 格構造関係

250 変換単語

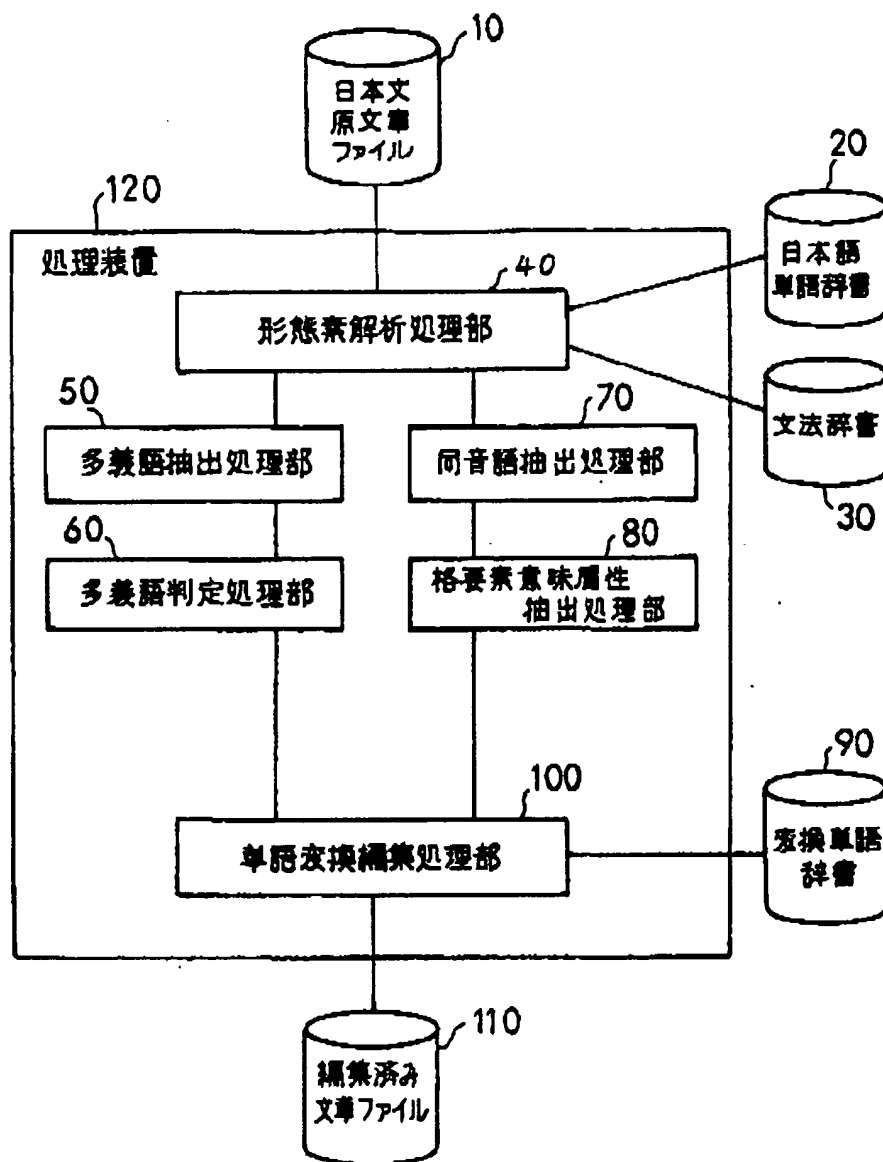
260 変換後文字列

(6)

特開平5-224687

【図1】

本発明の処理ブロック構成例



(7)

特開平5-224687

【図2】

日本語単語辞書の構成例

単語見出し	品 詞	読み	画形多値 フラグ	国音 フラグ	意味属性
日中	時制	ニッチュウ	1	0	時間
日中	固有名称	ニッチュウ	1	0	国、地域
取る	動詞	トル	0	1	
採る	動詞	トル	0	1	
摂る	動詞	トル	0	1	
撮る	動詞	トル	0	1	

20

【図3】

変換単語辞書の構成例

単語見出し	品 詞	意味属性		変換単語
		格要素単語	該当属性	
日中	時制		時間	変換の国
日中	固有名称		国、地域	日本と中国
取る	動詞	具体物		取る
取る	動詞	抽象物		取る
採る	動詞	割合		採用する
採る	動詞	次第		採用する
摂る	動詞	生物		採用する
撮る	動詞	生物		撮影する
撮る	動詞	抽象物		撮影する
撮る	動詞	具体物		撮影する

140

130

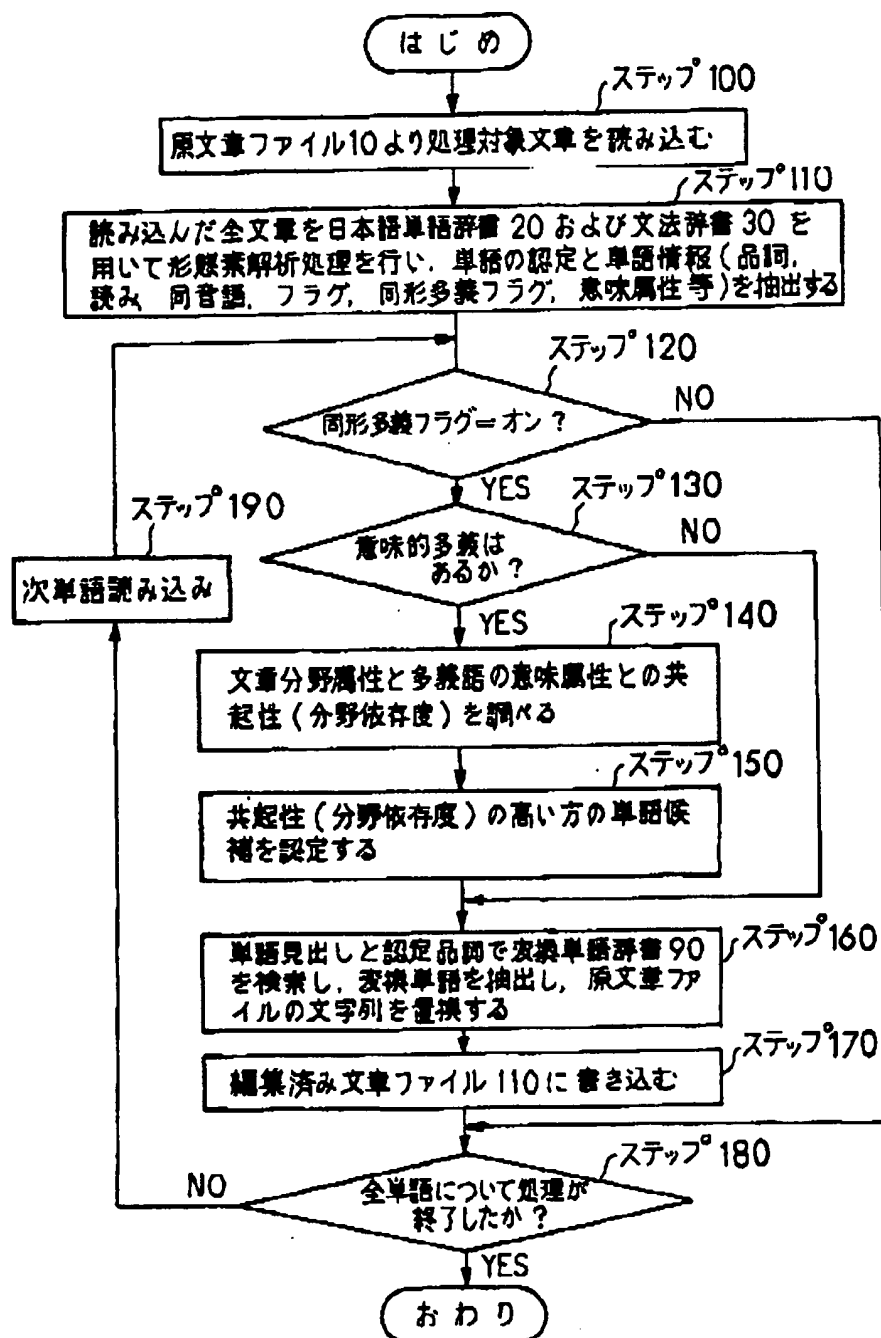
90

(8)

特開平5-224687

【図4】

多義語変換の処理概略フロー

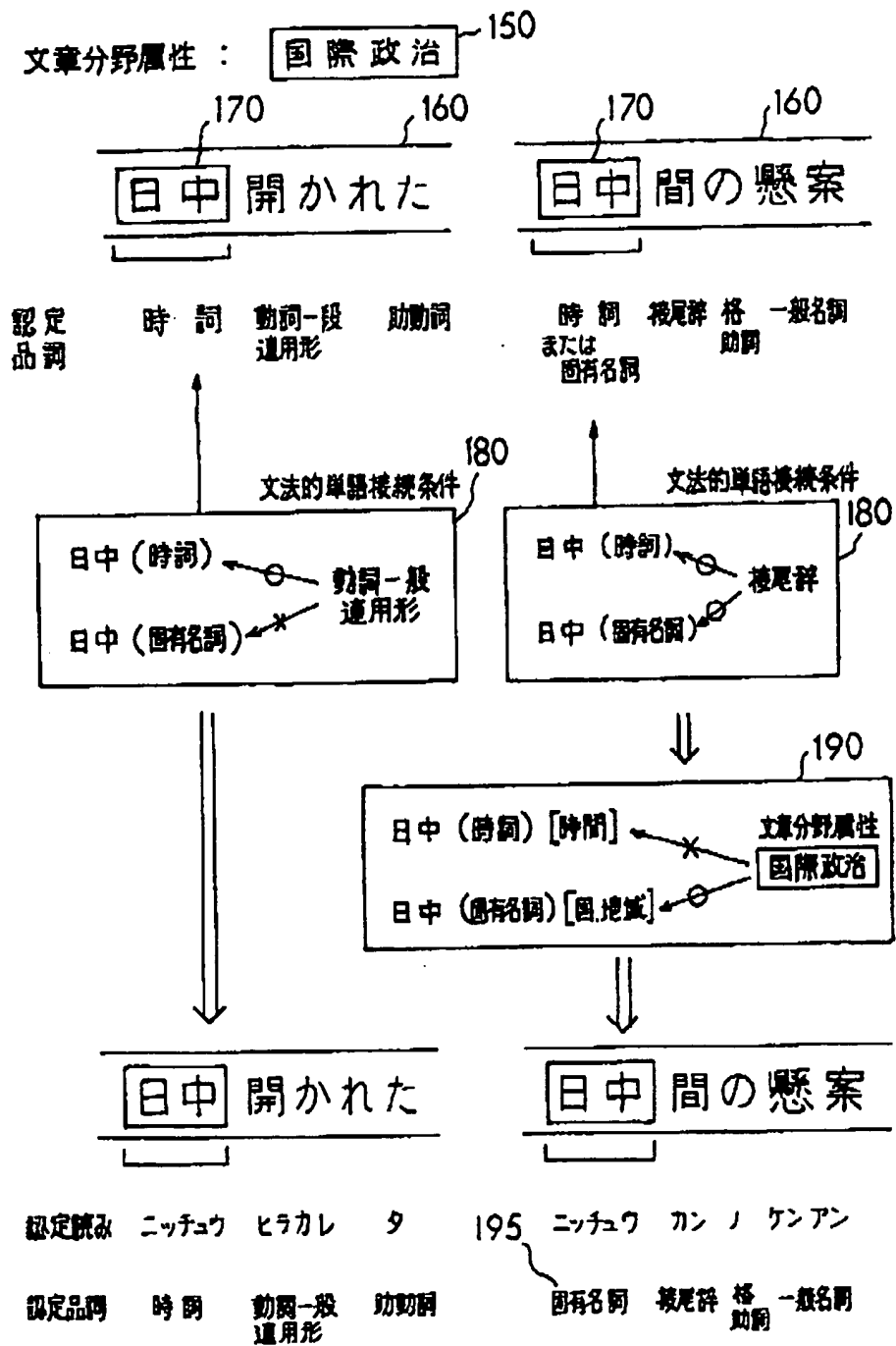


(9)

特開平5-224687

【図5】

多義語の多義判定処理

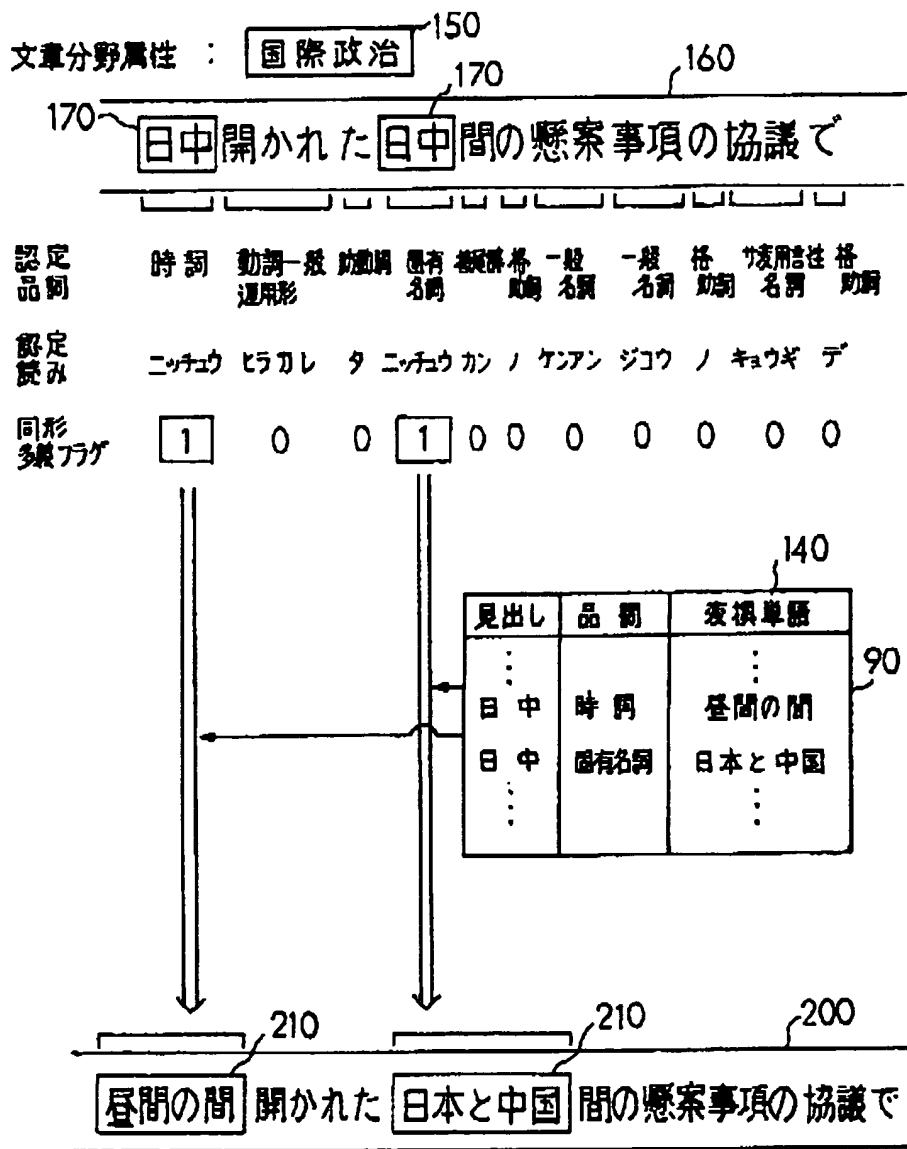


(10)

特開平5-224687

【図6】

多義語の単語変換処理



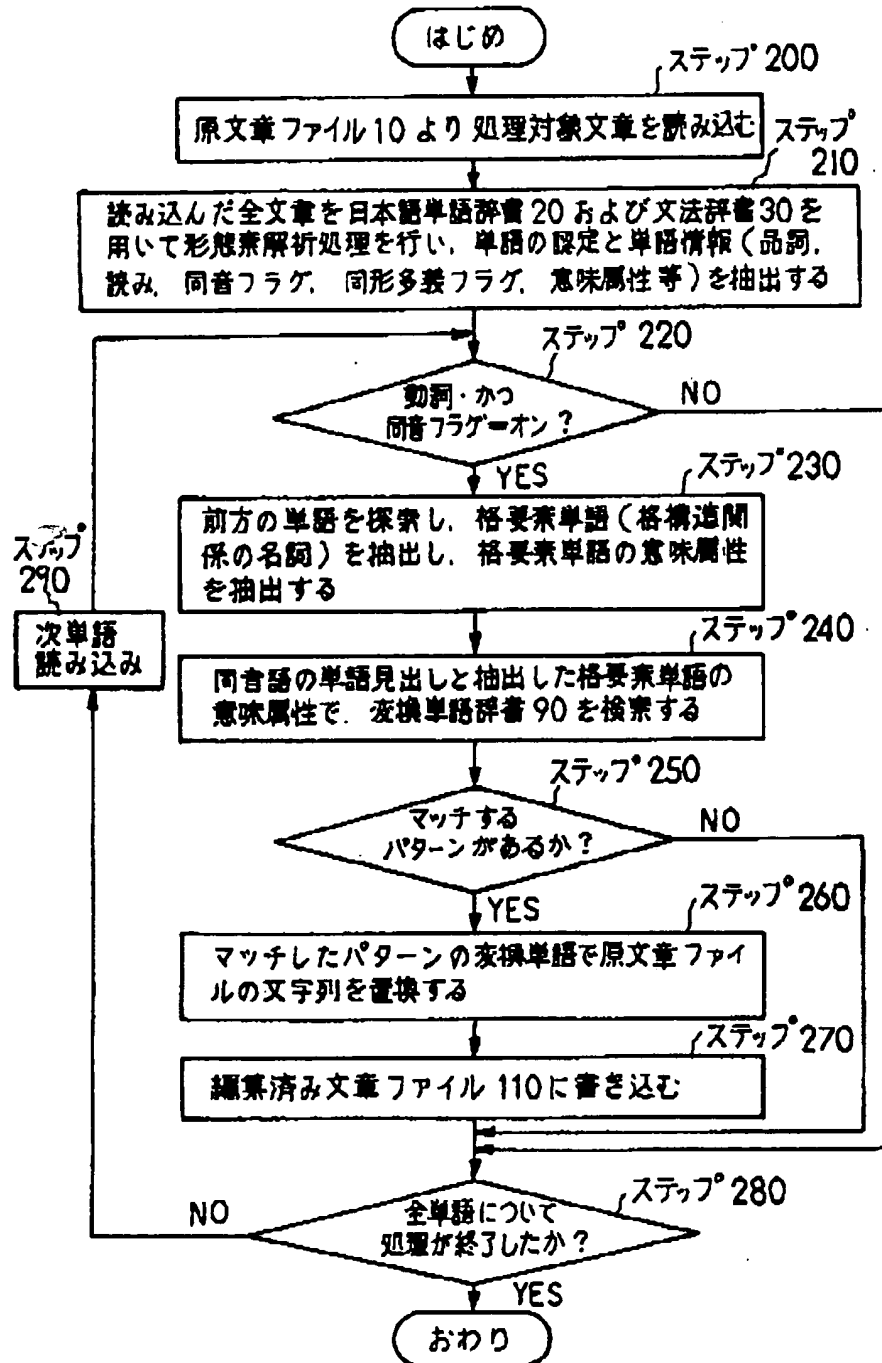
読み ヒルマノ アイダ ヒラカレ タ ニホントチュウグク カンノケンアンジコウノキョウギデ

(11)

特開平5-224687

【図7】

同音語変換の処理概略フロー



(1 2)

特開平5-224687

【图8】

同音語の変換編集処理

